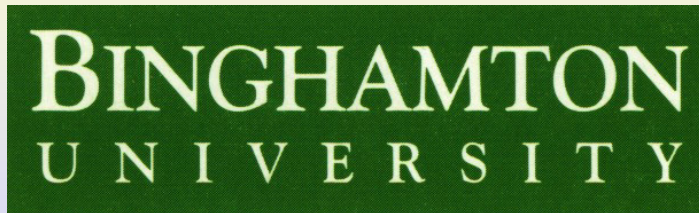


Breaking the OutGuess

Jessica Fridrich, Miroslav Goljan, Dorin Hoge^{}*
presented by Deepa Kundur

Department of Electrical and Computer Engineering
^{*}Department of Computer Science
SUNY Binghamton, Binghamton, NY 13902-6000, U.S.A,



State University of New York

Outline

- Introduction to steganography and steganalysis
- OutGuess steganographic algorithm
- Detection algorithm
- Experimental results
- Limitations and countermeasures
- Conclusion

Steganography (brief introduction)

The main goal of steganography is to hide the very presence of communication, such as by hiding messages in digital images

The most important requirement is that the act of embedding should not create any *statistically detectable artifacts* in stego images

It is not typically required that the data is embedded in a robust manner. Steganography is fundamentally different from watermarking.

Steganalysis

Steganalysis is the art of discovering the presence of secret data

Steganography has been broken if we can distinguish innocuous images from stego images with a success better than random guessing even though we may not be able to recover the embedded data

The goal of this paper is not only to distinguish cover images from stego images, but also to obtain an estimate of the length of the hidden message

OutGuess (part I)

Proposed by Neils Provos in 2001 as a response to the statistical chi-square attack by Andreas Westfeld in 1999

Main features of OutGuess:

- OutGuess hides messages in JPEG files
- It embeds message bits in LSBs of quantized DCT coefficients along a key-dependent walk through the image
- OutGuess preserves the histogram of DCT coefficients exactly
- OutGuess cannot be detected using the chi-square attack or its generalized versions

OutGuess (part II)

OutGuess works in two phases – embedding and correction steps.

Embedding

Original DCT coefficients	-4	1 skipped	3	0 skipped	1 skipped	8
Message bits	1	0	0	1	1	0
Modified DCT coefficients	-3	1 skipped	2	0 skipped	1 skipped	8

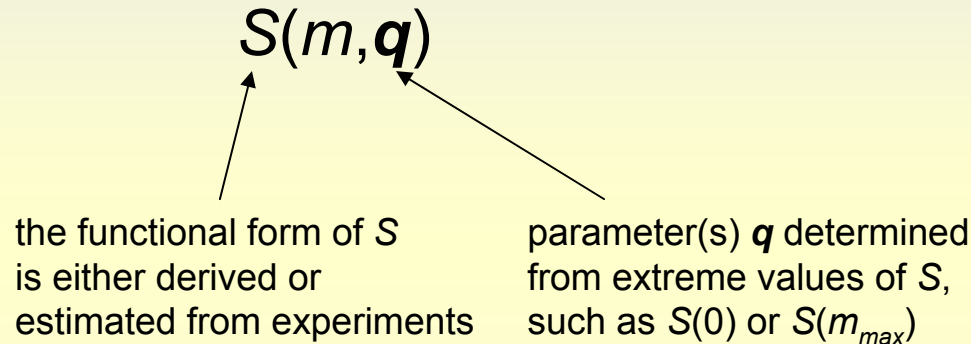
The embedding process skips 0's and 1's and flips the LSBs of coefficients to match them with the message bits

Correction

Because the embedding process changes the histogram of the quantized DCT coefficients, the correction steps flips LSBs of yet not visited DCTs to match the cover and stego histograms

Detection principle

We identify a macroscopic quantity $S(m, \mathbf{q})$ (distinguishing statistics) that predictably changes (for example, monotonically increases) with the length of the embedded secret message m . S depends on parameters \mathbf{q} .



Once the parameters have been determined, one can calculate an estimate of the unknown message length m by solving the equation

$$S(m) = S_{stego} \text{ for } m,$$

where S_{stego} is the value of S for the stego image under investigation.

Distinguishing statistics S

$$B = \sum_{i=1}^{\lfloor M-1/8 \rfloor} \sum_{j=1}^N |g_{8i,j} - g_{8i+1,j}| + \sum_{j=1}^{\lfloor N-1/8 \rfloor} \sum_{i=1}^M |g_{i,8j} - g_{i,8j+1}|$$

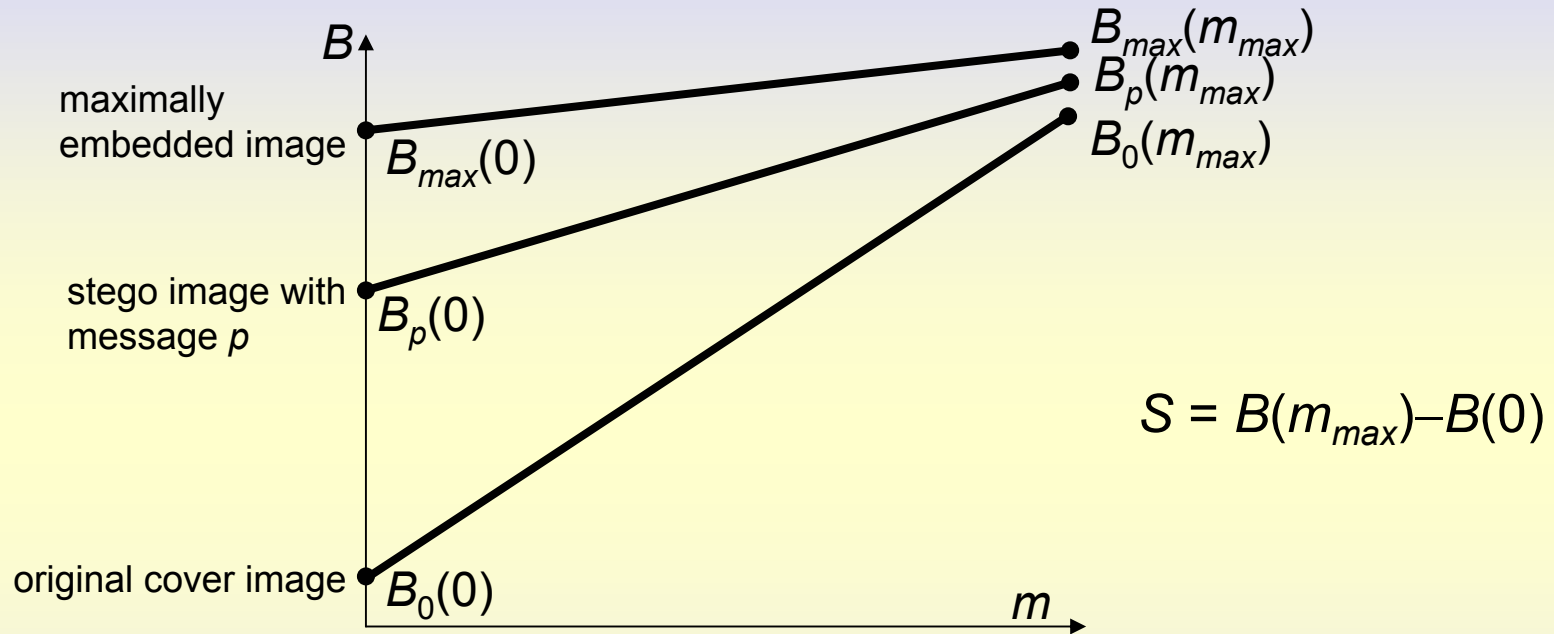
“Blockiness” B is the sum of spatial discontinuities along the boundary of 8×8 JPEG blocks.

B linearly increases with m – the number of bits embedded using OutGuess (experimental but well verified fact)

$$S = B(m_{max}) - B(0)$$

will be taken as our distinguishing statistics S .

S predictably changes with m



Because OutGuess uses LSB flipping as the main embedding mechanism, embedding another message into the stego image partially “cancels out”. Thus, S is largest for the cover image, smallest for the maximally embedded image, and somewhere “in-between” for a partially embedded image:

$$B_0(m_{max}) - B_0(0) > B_p(m_{max}) - B_p(0) > B_{max}(m_{max}) - B_{max}(0)$$

S_0 S_p S_{max}

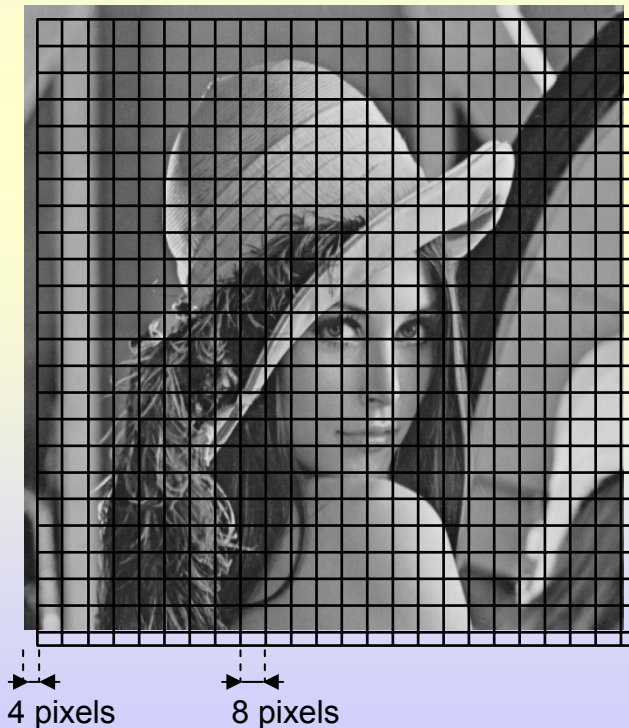
Detection algorithm

$S_p = B_p(m_{max}) - B_p(0)$ is known from the stego image

S_0 and S_{max} can be estimated from the stego image by cropping it and recompressing using the same quantization matrix.

The cropping and recompression breaks the quantization structure of DCT coefficients.

Because the cropped/recompressed image is perceptually close to the cover image, most macroscopic quantities, such as S_0 and S_{max} , are approximately preserved.



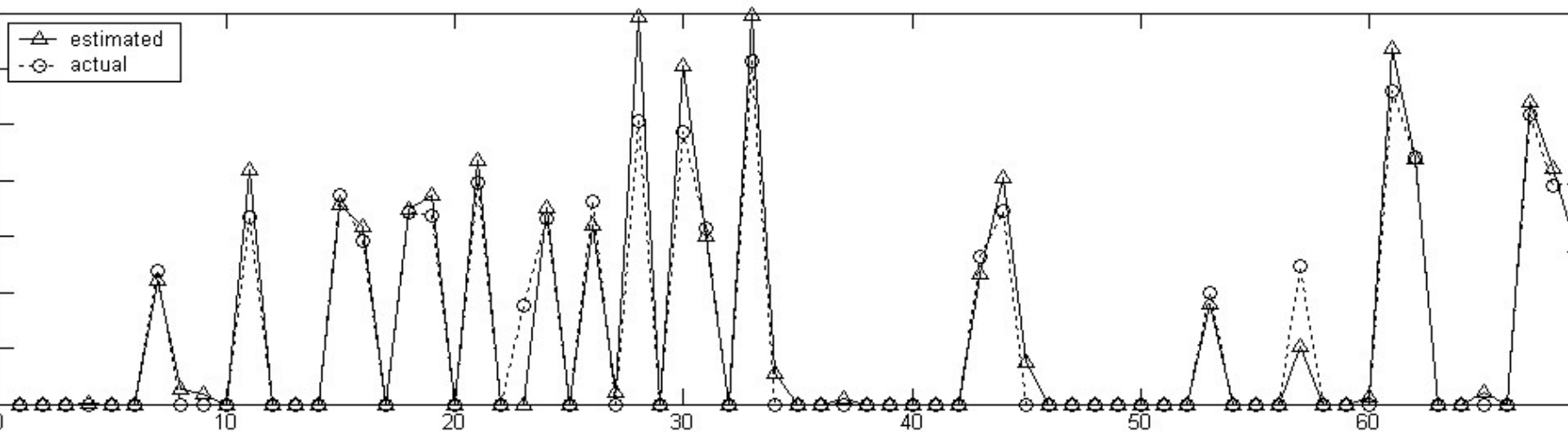
Detection algorithm

Because S is a linear function of the message length, the unknown message length p can be calculated as

$$p = \frac{S_0 - S}{S_0 - S_{\max}}$$

Experimental results

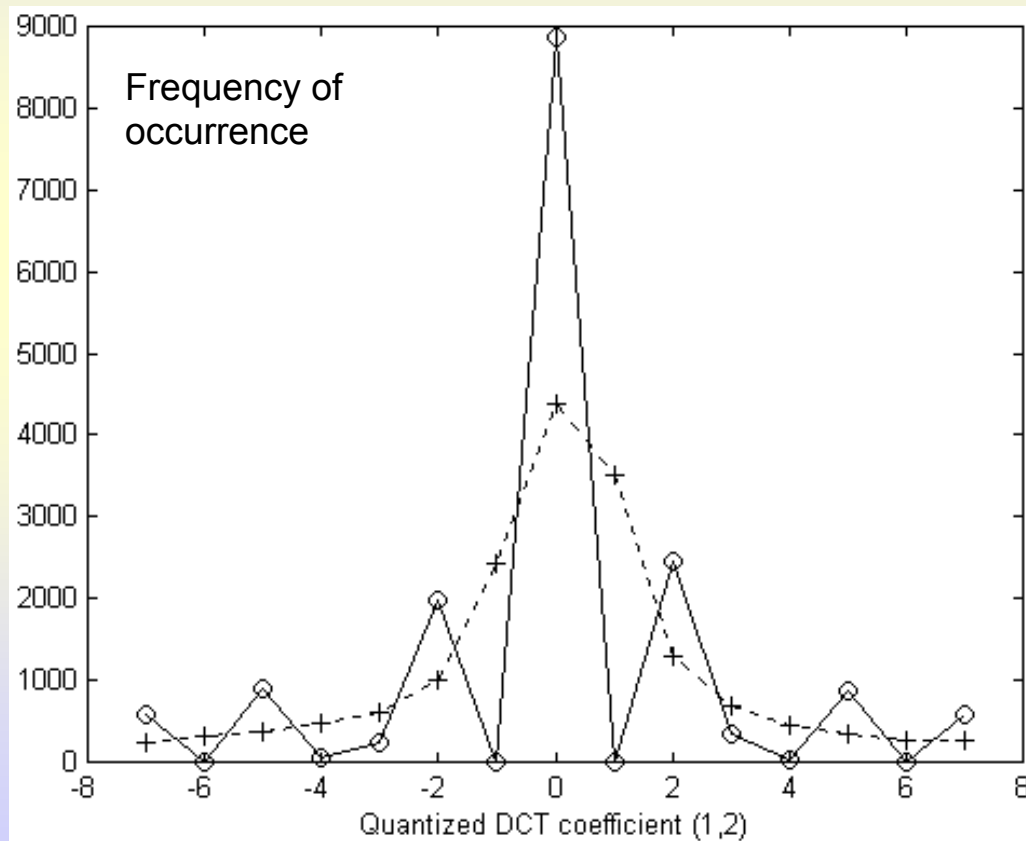
70 grayscale JPEG images compressed using quality factors ranging from 70 to 90 were used to test the detection routine.



y-axis: relative number of changes due to embedding (includes embedding and correction modifications)

Problem with double-compressed images

If a JPEG file is sent to OutGuess, it decompresses it first, then recompresses using a user-defined quality factor, and then embeds the message. This double compression complicates detection.



---+---+--- cover image

—o—o— double compressed cover image

Double-compression correction

- Double compression must be corrected for, otherwise a large error in message length estimation may result
- The cropped image can be used to estimate the primary quantization matrix of the cover image:

Q_s = quality factor of the stego image

h = histogram of DCT coefficients for the stego image

X_c = Cropped stego image

for $Q = 60 \dots 95$

Compress X_c using quality factor Q

Recompress it using Q_s and denote X_Q

h_c = histogram of X_Q

Calculate distance between histograms $d(Q) = \| h - h_c \|^2$

end

The primary quality factor $Q_s = \arg \min_Q d(Q)$

Lessons learned

By cropping and recompressing the stego image, we obtain a new JPEG file with many macroscopic properties close to the cover image

Thus, secure steganography must preserve all statistical measures that exhibit approximate invariance to cropping/recompressing, which might be a highly non-trivial task. It will also limit the already low capacity of JPEGs

The same approach can be used to attack F5:

J. Fridrich, M. Goljan, and D. Hoge, Steganalysis of JPEG Images: Breaking the F5 Algorithm, 5th Information Hiding Workshop, Noordwijkerhout, The Netherlands, October 2002